Supplementary Appendix 1

Fuller description of Materials and Methods

The study population was children born and diagnosed with cancer or a non-malignant brain tumour in Great Britain between 1980 and 2006, recorded on the National Registry of Childhood Tumours (NRCT) [1], an essentially complete population-based registry of cancers diagnosed in Great Britain before the child's fifteenth birthday. Birth registration details were available for almost all births in Britain. During this period, the annual incidence rate of childhood (0-14 years of age) cancer was around 140 per million; leukaemia, of which only about 2% are chronic, accounts for just under one-third of childhood cancer incidence [1].

A control matched on sex and date of birth (to within six months) had been previously selected for each NRCT case from the same birth register as the case; for cases diagnosed in 2000 and later years a second control had been selected in a similar way. The availability of residential address data and the numbers of controls were dictated by reasons unrelated to the present study. Controls were cancer-free at the time of diagnosis of their matched control. A total of 27 447 cases and 36 793 controls resulted. The principal outcome of interest in this study was leukaemia, but all types of childhood cancer were considered. Stringent diagnostic criteria are standard for NRCT [1] and conform to the third International Classification of Childhood Cancer (ICCC-3) [2], itself based on the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3) [3].

Residential addresses of mothers at the time of the birth of their child were assigned grid references using the Code-Point® system for all study participants [4]. In most instances (>96% of

records overall with very similar proportions for cases and controls) the more precise ADDRESS-POINT® [5] was known. The mean separation between the ADDRESS-POINT grid references and the centroid of the postcode (used by the Code-Point system) was 90 m for cases and 91m for controls. The ADDRESS-POINT grid reference of the maternal residence at birth was within 170 m of the postcode centroid for 95% of cases and controls. Full residential histories are not known for NRCT cases and controls, but addresses at diagnosis are available for the former.

Ionising radiation exposures of cases and controls were estimated from the place of residence of the mother at the child's birth. Data on indoor absorbed dose-rates from gamma-rays together with the directly-ionising component of cosmic rays came from the UK National Survey of natural background radiation based on six-month measurements made in 2333 houses, of which 2283 were in Great Britain (i.e. in England, Scotland and Wales)[6]; for brevity we denote these combined doses as "gamma-ray dose". For this study we used mean gamma-ray dose-rates in 459 English County Districts (CDs) and comparable administrative areas in Wales and Scotland derived from the data of the National Survey [7]. The average number of measurements per CD was 5.0 with a range from 0 to 35. In 19 CDs without measurements (including 304 cancer cases and 429 controls) the mean gamma-ray dose-rate for the county including the CD was used. If a case and its matched control(s) have residential addresses in the same CD then they will be assigned the same gamma-ray dose-rate. However, the measurement density did not permit gamma-ray estimates to be made for areas smaller than CDs. Gamma-ray dose-rates are generally more stable with time and less variable with geography than are radon concentrations, for which higher resolution data are available [6].

Two sources of radon concentration estimates for the homes of study subjects were used:

1. Mean exposures in CDs, from the National Survey [6]; the radon analogues of the gamma-ray estimates.

2. A predictive radon map based on the results of about 400 000 measurements of radon concentrations in homes grouped by grid squares and boundaries between different geological units [8 9 10]. This predictive map was developed by the Health Protection Agency and British Geological Survey (HPA/BGS). The results were normalised to the National Survey [6]. The most detailed radon mapping is based on geology and location specified by ADDRESS-POINT (GridSquare/AP estimates). For about 10% of children mapping based on the less precise Code-Point location was used.

Other components of natural background radiation (in particular the ingestion of naturally occurring radionuclides in food and drink) contribute almost as much to the RBM equivalent dose as do the components considered here [11], but the assessment of individual doses from these sources for the subjects of this study is not possible.

The National Survey reported gamma-ray absorbed dose-rate (in units of nGy/h) and radon activity concentration (Bq/m^3). Our main analyses use gamma-ray cumulative absorbed dose (mGy) and radon time-integrated exposure (kBq/m^3 years) for the period from the date of birth to the date of diagnosis (for controls, to the date of diagnosis of the corresponding case), which approximates the period from conception to nine months before diagnosis, i.e. since we take account of exposure in utero, the main analyses assume a minimum latent period of nine months. We also investigate other minimum latent periods of 0, 12 and 24 months, defined as the periods from birth to diagnosis:

plus 9 months (approximating conception to diagnosis),

minus 3 months (approximating conception to 12 months before diagnosis), and

minus 15 months (approximating conception to 24 months before diagnosis).

The quantities used in the analysis (cumulative gamma-ray dose and time-integrated radon concentration) are proportional to tissue doses from the two components separately. To compare the risk estimates from this study with published estimates, it was necessary to assess doses to the target tissue in question, and if the risks from gamma-rays and radon are to be examined together doses from both sources must be calculated on the same basis. This could be done only for leukaemia, for which the relevant quantity is the red bone marrow (RBM) equivalent dose (mSv) [12].

The conversion factor for gamma-rays is determined largely by shielding of the RBM by the body. The calculation of dose to the RBM from radon is less simple. There are two contributions, from radon gas and from the short-lived radioactive decay products. Radon gas has a four-day half-life and can be taken to be in equilibrium amongst body tissues, with those tissues having a relatively high fat content having somewhat higher concentrations. Assumptions must be made about the rate at which the short-lived decay products are taken up from the lung by body fluids, about their deposition in tissues and their subsequent loss by physical or biological processes; more details are given by Kendall et al [12]. Gamma-ray RBM doses during the fetal period are not greatly different from those after birth while radon doses are substantially smaller. However, for the present work, conversion factors from the measured quantities to RBM equivalent dose are based on the mean dose from birth to age 15 years. The conversion factors to RBM equivalent dose (mSv) that result are: for gamma-ray dose (mGy), 0.79, and for radon time-integrated activity concentrations (kBq m^{-3} – years), 3.4.

Socio-economic status (SES) is known to influence rates of childhood cancer, particularly leukaemia [13 14]. For the SES quantities used here leukaemia incidence rates are normally higher in more affluent groups, although there is evidence that this might not apply to other SES quantities. [15] The principal measure of SES considered in the analysis was the Carstairs deprivation index, based upon the census ward in which the mother was living at the child's birth [14, 16]; the main analysis included quintiles of the Carstairs index as a measure of deprivation. An alternative measure of SES was the social class of the father, derived from his occupation as stated on the child's birth record. The occupational description was coded and social class category derived from classifications used by the Office of Population Censuses and Surveys, now the Office for National Statistics [17 18]. Social class derived in this way was not available for all study members and was based on self-reported data, which were sometimes ambiguous.

Carstairs scores were available for all cases and controls in the study. Paternal occupational social class was available for about 90% of cases and controls, but analyses including this variable were restricted to the 85% of matched sets where both case and control(s) were assigned a value.

Statistical Methods

The analysis used conditional logistic regression [19] implemented in STATA [20]. The probability of developing cancer for individual j in stratum i (as given by an indicator variable $Y_{i,j} = 1$ if cancer, and $Y_{i,j} = 0$ if not) with cumulative lagged dose $D_{i,j}$ (lagged by 9 months in the main analysis, but using also 0, 12 and 24 months in subsidiary analyses) and Carstairs score $S_{i,j}$ is given by the standard logistic model:

$$P[Y_{i,j} \mid D_{i,j}, S_{i,j}] = \frac{\exp\left[\alpha_{0,i} + \alpha_1 D_{i,j} + \alpha_2 S_{i,j}\right]^{Y_{i,j}}}{1 + \exp\left[\alpha_{0,i} + \alpha_1 D_{i,j} + \alpha_2 S_{i,j}\right]} \quad \text{(A1)}$$

Therefore, the odds ratio for the individual j relative to individual j_0 is given by:

$$\frac{\frac{P[Y_{i,j}=1 \mid D_{i,j}, S_{i,j}]}{P[Y_{i,j}=0 \mid D_{i,j}, S_{i,j}]}}{\frac{P[Y_{i,j_0}=1 \mid D_{i,j_0}, S_{i,j_0}]}{P[Y_{i,j_0}=0 \mid D_{i,j_0}, S_{i,j_0}]}} = \exp\left[\alpha_1[D_{i,j} - D_{i,j_0}] + \alpha_2[S_{i,j} - S_{i,j_0}]\right] \quad \text{(A2)}$$

Then the conditional probability of individual $j = 0$ being the case and individuals $j = 1,...,K_i$ being the controls is:

$$P[\, j = 0 \text{ is case}, j = 1,...,K_i \text{ are controls} | D_{i,j}, S_{i,j}, j = 0,...,K_i] \quad \text{(A3)}$$

$$= \frac{P\left[Y_{i,0}=1 \mid D_{i,0}, S_{i,0}\right] \prod_{j=1}^{K_i} P\left[Y_{i,j}=0 \mid D_{i,j}, S_{i,j}\right]}{\sum_{j=0}^{K_i} P\left[Y_{i,j}=1 \mid D_{i,j}, S_{i,j}\right] \prod_{m \neq j} P\left[Y_{i,m}=0 \mid D_{i,m}, S_{i,m}\right]}$$

$$= \frac{\exp\left[\alpha_1 D_{i,0} + \alpha_2 S_{i,0}\right]}{\sum_{j=0}^{K_i} \exp\left[\alpha_1 D_{i,j} + \alpha_2 S_{i,j}\right]}$$

In the present study the number of controls is always $K_i = 1$ or $K_i = 2$. The conditional likelihood is simply the product over all matched case-control sets of these terms. This model was fitted via maximum likelihood [21].

As in (A2), the odds ratio (which is very close to the relative risk (RR) when the probability of disease is (as here) low [19]) is given by $OR = \exp\left[\alpha_1[D_{i,j} - D_{i,j_0}] + \alpha_2[S_{i,j} - S_{i,j_0}]\right]$. With a slight abuse of notation, we generally present results as RRs relative to the zero dose

group, so that $RR = \exp\left[\alpha_1 D_{i,j}\right]$. Confidence intervals (CI) were Wald-based, calculated using the Fisher information [22]. The p-values presented were calculated from likelihood-ratio tests, and are two-sided. Heterogeneity across strata was assessed by considering the value of the deviance difference statistic (in relation to the chi-squared distribution) with the appropriate number of degrees of freedom for combining the RRs for radon and gamma individually. The "observed" dose-response plotted in the figure used a semi-parametric model with separate evaluation of risk over 0-, 0.5-, 1.0-, …, 7.5-, 8.0-, 9.0-, 10.0-, 11.0-, 12.0- mGy, and a 5-point moving average over neighbouring dose intervals (with weights (0.15, 0.2, 0.3, 0.2, 0.15) and the inverse variance of each point) used to derive a smoothed spline.

The reporting of this study conforms to the STROBE statement [23].

Reference List

1 Stiller C.Stiller C (editor) *Childhood cancer in Britain: incidence, survival, mortality*. Oxford University Press: Oxford, 2007.

2 Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer* 2005; **103**: 1457-1467.

3 World Health Organisation.Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM et al. (eds) *International Classification of Diseases for Oncology*. 3rd edition ed. WHO: Geneva, 2000.

4 Royal Mail. *Postcode Atlas of Great Britain and Northern Ireland*. Bartholomew: London, 1996.

5 Martin D, Higgs G. Population georeferencing in England and Wales: basic spatial units reconsidered. *Environ Plan A* 1997; **29**: 333-347.

6 Wrixon AD, Green BMR, Lomas PR, Miles JCH, Cliff KD, Francis EA *et al*. Natural radiation exposure in UK dwellings. National Radiological Protection Board: Chilton,Didcot,Oxon, 1998. Report No: NRPB-R190.

7 Kendall GM, Hughes JS, Oatway WB, Jones AL. Variations in radiation exposures of adults and children in the UK. *J Radiol Prot* 2006; **26**: 257-276.

8 Miles JCH, Appleton JD. Mapping variation in radon potential both between and within geological units. *J Radiol Prot* 2005; **25**: 257-276.

9 Miles JCH, Appleton JD, Rees DM, Green BMR, Adlam KAM, Myers AH. Indicative atlas of radon in England and Wales. Health Protection Agency: Chilton, 2007. Report No: HPA-RPD-033.

10 Appleton JD, Miles JCH. A statistical evaluation of the geogenic controls on indoor radon concentrations and radon risk. *J Environ Radioact* 2010; **101**: 799-803.

11 Barr RD, Greenberg ML. Cancer surveillance and control in adolescents - similarities and contrasts between Canada and the United States. *Pediatr Blood Cancer* 2006; **46**: 273-277.

12 Kendall GM, Fell TP, Harrison JD. Dose to red bone marrow of infants, children and adults from radiation of natural origin. *J Radiol Prot* 2009; **29**: 123-138.

13 Committee on Medical Aspects of Radiation in the Environment (COMARE). COMARE 11th report: The distribution of childhood leukaemia and other childhood cancer in Great Britain 1969-1993. Health Protection Agency: Chilton, Didcot, 2006.

14 Kroll ME, Stiller CA, Murphy MFG, Carpenter LM. Childhood leukaemia and socioeconomic status in England and Wales 1976–2005: evidence of higher incidence in relatively affluent communities persists over time. *Br J Cancer* 2011; **105**: 1783-1787.

15 Poole C, Greenland S, Luetters C, Kelsey JL, Mezei G. Socioeconomic status and childhood leukaemia: a review. *Int J Epidemiol* 2006; **35**: 370-384.

16 Carstairs V, Morris R. *Deprivation and Health in Scotland*. Aberdeen University Press: Aberdeen, 1991.

17 Office of Population Censuses & Surveys. *Classification of Occupations 1980*. HMSO: London, 1980.

18 ONS Classification and Harmonisation Unit. CO80 simplified 1980 occupation classification ONS factsheet. 1985. 1985. Crown Copyright.

19 Breslow NE, Day NE.Davis W (editor) *Statistical Methods in Cancer Research Volume 1 - The Analysis of Case-Control Studies IARC Scientific Publications No.32*. International Agency for Research on Cancer: Lyon, 1980.

20 StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.

21 McCullagh P, Nelder JA. *Generalised Linear Models*. Second edition ed. Chapman & Hall/CRC, 1989.

22 Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman & Hall: London, 1974.

23 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; **335**: 806-808.